# EDGEWORTH EXPANSIONS AND JACKKNIFE METHODS

David Hinkley* and Wei Bo-Cheng

School of Statistics, University of Minnesota

Technical Report No. 406

October 1982

Summary

Edgeworth expansion methods are applied to Studentized estimates where the standard error has been computed by a jackknife method. Consequential adjustments to the standard jackknife confidence limit technique are discussed. Results are illustrated numerically for estimation of the ratio of two means.

Key words: Jackknife, bootstrap, Edgeworth expansion, statistical functional, ratio estimate.

*Research supported in part by NSF Grants MCS-7904558, MCS-8201994.

## 1. INTRODUCTION

The standard jackknife method is commonly used as a nonparametric method for estimating bias and standard error of a statistical estimate T. An associated confidence limit method is then determined by use of a normal approximation for T, often justified by appeal to asymptotic theory for large sample size. Somewhat curiously, there seems to be little discussion of improvements to the normal approximation in this context, although such improvements are familiar in connection with the sample average, Student's t statistic, linear rank statistics, etc. Standard techniques used in these latter cases are those associated with Edgeworth expansions. The present paper discusses how such expansions can be implemented in the jackknife context, and a simple example is discussed.

We consider the following situation: $X_1,\ldots,X_n$ are independent homogeneous variables with common c.d.f. F, one characteristic of which is the single parameter $\theta = t(F)$; for example, the mean is $\mu = m(F) = \int x\,dF(x)$. We wish to obtain confidence limits for $\theta$ based on the nonparametric m.l.e. $T_n = t(\hat{F}_n)$, where $\hat{F}_n$ is the sample c.d.f.

$$\hat{F}_n(x) = n^{-1}\sum_{k=1}^{n}\Delta(x-X_k) = \frac{\#\, X_k \le x}{n} \quad ; \tag{1.1}$$

here $\Delta(z) = 0$ or $1$ according as $z < 0$ or $z \ge 0$. Specific confidence limit methods are based on Studentized quantities of the form $(T_n - \theta)/SE_n$, where the standard error $SE_n$ is computed by a jackknife technique.

There are several forms of jackknife, but here we consider only two: the <u>standard</u> (Quenouille-Tukey) jackknife and the <u>infinitesimal</u> (Jaeckel) jackknife. These can be described briefly as follows

<u>Standard jackknife</u>. First compute the quantities

$$(1.2) \qquad \tilde{I}_j = (n-1)(T_n - T_{n/j}) = (n-1)\{t(\hat{F}_n) - t(\hat{F}_{n/j})\} , \quad j=1,\ldots,n ,$$

where subscript n/j denotes computation from the sample without $X_j$, i.e. $\{X_k : 1 \leq k \neq j \leq n\}$. Next calculate

$$(1.3) \qquad \tilde{B}_n = -\bar{\tilde{I}}_{\bullet} = -n^{-1}\Sigma\tilde{I}_j , \quad \tilde{V}_n = \frac{1}{n-1}\Sigma(\tilde{I}_j - \bar{\tilde{I}}_{\bullet})^2$$

as estimates of bias $B_n = E(T_n) - \theta$ and $V_n = n\mathrm{Var}(T_n)$, respectively. Then the Studentizing quantity $SE_n$ is taken to be $\sqrt{n^{-1}\tilde{V}_n}$, and either $\sqrt{n}(T_n-\theta)/\sqrt{\tilde{V}_n}$ or $\sqrt{n}(T_n-\tilde{B}_n-\theta)/\sqrt{\tilde{V}_n}$ is taken to be approximately standard normal. Thus, for example, approximate 95% confidence limits for $\theta$ would be

$$(1.4) \qquad T_n \pm 2\sqrt{\tilde{V}_n/n} \qquad \text{or} \qquad T_n - \tilde{B}_n \pm 2\sqrt{\tilde{V}_n/n} .$$

<u>Infinitesimal jackknife</u>. First compute the quantities

$$(1.5) \qquad \hat{I}_j = I_t(X_j,\hat{F}_n) , \qquad j=1,\ldots,n ,$$

where $I_t(x,F)$ is the influence function of $t(\cdot)$ at $(x,F)$; see Section 2. Next, calculate

$$(1.6) \qquad \hat{V}_n = n^{-1}\,\Sigma\,\hat{I}_j^2$$

to estimate $V_n = n\mathrm{Var}(T_n)$, and take $\sqrt{n}(T_n-\theta)/\sqrt{\hat{V}_n}$ to be approximately standard normal.

We should note that in the latter method a bias estimate can be calculated (see Section 2). However, the bias is of smaller order than the standard error, so that its inclusion in a first-order method such as (1.4) gives false security: other terms of comparable magnitude, such as skewness, should also be taken into account if the first-order theory provides an inadequate approximation. In what follows, bias and skewness of $T_n$ and variation of standard error are all seen to contribute to second-order approximations. This is more complicated than the situation for the simple statistic $\sqrt{n}(\overline{X}-\mu)$, but standard Edgeworth expansion techniques still apply.

The main focus of the paper is on the infinitesimal jackknife, because the estimate $\hat{V}_n$ in (1.6) is a simple explicit functional of $\hat{F}_n$, whereas $\tilde{V}_n$ in (1.3) is not. Section 2 describes Edgeworth expansions for

$$\hat{Z}_n = \sqrt{n}(T_n-\theta)/\sqrt{\hat{V}_n} \tag{1.7}$$

and establishes a first-order correction $\hat{a}_\rho$ to the standard normal $\rho$-quantile $k_\rho^N$ such that

$$\Pr(\hat{Z}_n \le k_\rho^N + \hat{a}_\rho) = \rho + O(n^{-1}) \quad .$$

It turns out (Section 3) that this last result applies also to the standard jackknife. Section 4 evaluates the usefulness of the results in the context of ratio estimation, and compares jackknife methods to bootstrap methods.

## 2. EDGEWORTH EXPANSIONS FOR INFINITESIMAL JACKKNIFE

### 2.1 Von Mises Expansion of Studentized Statistic

As described in Section 1, the infinitesimal jackknife is based on the empirical influence function $I_t(x,\hat{F}_n)$, specifically on the variance estimate

$$\hat{V}_n = v(\hat{F}_n) = \int I_t^2(x,\hat{F}_n)d\hat{F}_n(x) \quad . \tag{2.1}$$

A formal definition of $I_t$ is

$$I_t(x,F) = \frac{d}{d\varepsilon}\, t(F_{\varepsilon,x})\big|_{\varepsilon=0} \tag{2.2}$$

where $F_{\varepsilon,x}(y) = (1-\varepsilon)F(y) + \varepsilon\Delta(y-x)$. Note that in practice (2.2) can be approximated by a numerical difference for small $\varepsilon$. Thus $\hat{I}_j = I_t(X_j,\hat{F}_n)$ could be approximated by

$$\varepsilon^{-1}\{t(\hat{F}_{n,\varepsilon,X_j}) - t(\hat{F}_n)\} \qquad ;$$

$t(\hat{F}_{n,\varepsilon,X_j})$ corresponds to $T_n$ but calculated for a sample with relative frequencies $(1-\varepsilon)/n$ at $X_k (k\neq j)$ and $\varepsilon + (1-\varepsilon)/n$ at $X_j$. One might choose $\varepsilon = 10^{-6}$, certainly $\varepsilon \ll n^{-1}$; the standard jackknife is equivalent to the choice $\varepsilon = -(n-1)^{-1}$.

The Studentized form of $T_n$ is

$$(2.3) \qquad \hat{Z}_n = \frac{\sqrt{n}(T_n-\theta)}{\sqrt{\hat{V}_n}} = \frac{\sqrt{n}\{t(\hat{F}_n)-t(F)\}}{\sqrt{v(\hat{F}_n)}} = \sqrt{n}\,\omega(\hat{F}_n,F) \quad ,$$

say, where $v(\cdot)$ is defined in (2.1). The large-sample standard normal approximation for $\hat{Z}_n$ is justified by weak convergence of $v(\hat{F}_n)$ to $v(F)$ and a central limit theorem for a linear approximation to $t(\hat{F}_n)-t(F)$. Our interest is in going further, which will involve a series expansion for $\omega(\hat{F}_n,F)$. It is useful to begin with $t(\hat{F}_n)-t(F)$.

The von Mises expansion of $t(\cdot)$ is

$$2.4) \qquad t(G) = t(F) + \sum_{k=1}^{r} \int \cdots \int D^k t(x_1,\ldots,x_k,F)\,\frac{1}{k!}\prod_{i=1}^{k} d(G-F)(x_j) + \rho_r \quad ,$$

where $\rho_r = O(||G-F||^r)$ with, e.g., $||G-F|| = \sup|G(x)-F(x)|$. The term $D^k t$ in (2.4) is the k-th functional derivative of t, and is assumed symmetric in the $x_j$'s. For our purposes it is convenient to make the derivatives unique by requiring that <u>all full and partial expectations of</u> $D^k t(X_1,\ldots,X_k,F)$ <u>are zero for independent</u> $X_i$'s <u>each having c.d.f.</u> F. Then we can re-express (2.4) as follows, now for the particular case $r=3$,

$$(2.5)\qquad t(G) \doteq t(F) + \int I_t(x_1,F)dG(x_1) + \frac{1}{2}\iint Q_t(x_1,x_2,F)dG(x_1)dG(x_2) + \frac{1}{6}\iiint C_t(x_1,x_2,x_3,F)dG(x_1)dG(x_2)dG(x_3)\ ,$$

with the following specific definitions: $I_t$ as in (2.1), then sequentially

$$(2.6)\qquad Q_t(x_1,x_2,F) = I_t(x_2,F) + \text{first derivative of } I_t(x_1,F) \text{ at } (x_2,F)\ ,$$

$$C_t(x_1,x_2,x_3,F) = Q_t(x_1,x_3,F) + Q_t(x_2,x_3,F) + \text{first derivative of } Q_t(x_1,x_2,F) \text{ at } (x_3,F)\ .$$

Substitution of $G = \hat{F}_n$ in (2.5) gives an expansion of $t(\hat{F}_n)-t(F)$ with error $O_p(n^{-3/2})$ , since $\sup|\hat{F}_n(x)-F(x)| = O_p(n^{-1/2})$.

Returning now to the statistic of interest, $\sqrt{n}\,\omega(\hat{F}_n,F)$ as defined in (2.3), we have $\omega(G,F) = \{t(G)-t(F)\}/\sqrt{v(G)}$ . The second argument, F, of $\omega$ is fixed. Thus we apply (2.5) directly for $\omega(G,F) = \omega(G,F) - \omega(F,F)$ , and substitute $G = \hat{F}_n$ to obtain

$$(2.7)\qquad \omega(\hat{F}_n,F) \doteq \omega_3(\hat{F}_n,F) = n^{-1}\Sigma\, I_\omega(X_j,F) + \frac{1}{2}n^{-2}\Sigma\Sigma\, Q_\omega(X_j,X_k,F) + \frac{1}{6}n^{-3}\Sigma\Sigma\Sigma\, C_\omega(X_j,X_k,X_\ell,F)\ .$$

The required derivatives of $\omega$ are obtained from those for t, using the intermediate results

$$(2.8)\qquad \begin{aligned} I_v(x_1,F) &= I_t^2(x_1,F) - V(F) + 2\,E\,\{I_t(Y,F)Q_t(x_1,Y,F)\} \\ Q_v(x_1,x_2,F) &= 2\{I_t(x_1,F) + I_t(x_2,F)\}Q_t(x_1,x_2,F) - 2I_t(x_1,F)I_t(x_2,F) \\ &\quad + 2\,E\,[\{C_t(x_1,x_2,Y,F) - Q_t(x_1,Y,F) - Q_t(x_2,Y,F)\} \\ &\quad \times I_t(Y,F) + Q_t(x_1,Y,F)Q_t(x_2,Y,F)]\ ; \end{aligned}$$

see Appendix. In particular, we find

$$I_\omega(x_1,F) = V^{-1/2}\, I_t(x_1,F)$$

$$(2.9) \qquad Q_\omega(x_1,x_2,F) = V^{-1/2}Q_t(x_1,x_2,F) - \frac{1}{2}\, V^{-3/2}\, \mathrm{Sym}\{I_t(x_1,F)I_v(x_2,F)\}$$

$$C_\omega(x_1,x_2,x_3,F) = V^{-1/2}C_t(x_1,x_2,x_3,F) - \frac{1}{2}\, V^{-3/2}[\mathrm{Sym}\{I_t(x_1,F)Q_v(x_2,x_3,F)\} + \mathrm{Sym}\{I_v(x_1F)Q_t(x_2,x_3,F)\}] + \frac{3}{4}\, V^{-5/2}\mathrm{Sym}\{I_t(x_1,F)I_v(x_2,F)I_v(x_3,F)\}$$

where Sym stand for symmetrized sum, eg. $\mathrm{Sym}\{a(x_1)b(x_2)\} = a(x_1)b(x_2) + a(x_2)b(x_1)$.

## 2.2 Edgeworth Expansion and Confidence Limits

Given the approximation (2.7), we now appeal to a general form of Edgeworth expansion described by Withers (1983); for theoretical details see also Reeds (1976, Chapter 5), Bhattacharya and Ghosh (1978). Specifically, we have the form

$$(2.10) \qquad \Pr\{\sqrt{n}\,\omega(\hat{F}_n,F) \le z\} = \Phi(z) - \phi(z) \sum_{r=1}^{2} n^{-r/2}\lambda_r(z) + O(n^{-1}) \quad ,$$

where $\phi$ and $\Phi$ are the standard normal p.d.f. and c.d.f., respectively. The terms $\lambda_1,\lambda_2$ are derived from the cumulants of the three term approximation $\omega_3$ in (2.7) as follows: Write the a-th cumulant of $\omega_3(\hat{F}_n,F)$ as

$$\sum_{b=a-1}^{3} \alpha_{ab}\, n^{-b} \qquad , \quad a \le 4 \quad ;$$

note that $\alpha_{10} = 0$, $\alpha_{21} = 1$ because $\omega(\hat{F}_n,F)$ is standardized. Then

$$(2.11) \qquad \lambda_1(z) = \alpha_{11} + \frac{1}{6}\,\alpha_{32}(z^2-1)$$

$$(2.12) \qquad \lambda_2(z) = \frac{1}{2}(\alpha_{11}^2+\alpha_{22})z + \frac{1}{24}(4\,\alpha_{11}\alpha_{32}+\alpha_{43})(z^3-3z) + \frac{1}{72}\,\alpha_{32}^2(z^5-10\,z^3+9\,z) \quad .$$

Straightforward calculation shows the required coefficients $\alpha_{ab}$ to be

$$\alpha_{11} = \frac{1}{2} E\, Q_\omega(X,X,F)$$

(2.13)

$$\alpha_{32} = E\, I_\omega^3(X,F) + 3\, E\{I_\omega(X,F) I_\omega(Y,F) Q_\omega(X,Y,F)\}$$

for $\lambda_1$, and then, for $\lambda_2$,

$$\alpha_{22} = E\{I_\omega(X,F)Q_\omega(X,X,F)\} + \frac{1}{2}\, E\, Q_\omega^2(X,Y,F)$$

$$+ E\{I_\omega(X,F)C_\omega(X,Y,Y,F)\} \quad ,$$

(2.14)

$$\alpha_{43} = E\, I_\omega^4(X,F) - 3 + 12\, E\{I_\omega(X,F)I_\omega^2(Y,F)Q_\omega(X,Y,F)\}$$

$$+ 12\, E\{I_\omega(X,F)I_\omega(Y,F)Q_\omega(X,Z,F)Q_\omega(Y,Z,F)\}$$

$$+ 4\, E\{I_\omega(X,F)I_\omega(Y,F)I_\omega(Z,F)C_\omega(X,Y,Z,F)\} \; .$$

(In these expressions X,Y,Z are independent each with distribution F.)

The terms $\alpha_{11}$ and $\alpha_{22}$ correspond respectively to the (standardized) first-order biases of $T_n$ and $\hat{V}_n$, while $\alpha_{32}$ and $\alpha_{43}$ include skewness and kurtosis measures of the linear approximation $n^{-1}\Sigma I_t(X_j,F)$ for $T_n-\theta$. When this linear approximation is exact, i.e. $Q_t \equiv 0$, the formulae simplify considerably and reproduce the known result for Student's t statistic; see, e.g., Hall (1983).

Now consider the problem of calculating confidence limits for $\theta$. To obtain corrections to the normal approximation limits described in Section 1, i.e.,

(2.15) $$T_n - z_{1-\rho}^N \sqrt{\hat{V}_n/n} \; , \quad T_n - z_\rho^N \sqrt{\hat{V}_n/n} \quad ,$$

we need to invert the Edgeworth expansion (2.10) and use estimates for coefficients $\alpha_{ab}$. This is straightforward for a first-order correction, ignoring the $n^{-1}$ term in (2.10). For then, with $\hat{\alpha}_{ab} = \alpha_{ab}(\hat{F}_n)$, we have

$$\Pr\{\sqrt{n}\omega(\hat{F}_n,F) \leq z\} = \Phi(z) - n^{-1/2}\{\hat{\alpha}_{11} + \frac{1}{6}\hat{\alpha}_{32}(z^2-1)\}\phi(z) + O_p(n^{-1}) .$$

Consequently, defining

$$\hat{z}_\rho = z_\rho^N + n^{-1/2}[\hat{\alpha}_{11} + \frac{1}{6}\hat{\alpha}_{32}\{(z_\rho^N)^2 - 1\}] \quad , \tag{2.16}$$

we have $\Pr\{\sqrt{n}\omega(\hat{F}_n,F) \leq \hat{z}_\rho\} = \rho + O(n^{-1})$. The first-order adjustments to limits (2.15) are therefore, by (2.3),

$$T_n - \hat{z}_{1-\rho}\sqrt{\hat{V}_n/n} \ , \ T_n - \hat{z}_\rho \sqrt{\hat{V}_n/n} \quad ; \tag{2.17}$$

by symmetry $-\hat{z}_{1-\rho} = \hat{z}_\rho$.

One point to note is that the $n^{-1/2}$ adjustment on the right of (2.16) is an even function, so that the symmetric normal limits (2.15) cover $\theta$ with probability $1-2\rho+O(n^{-1})$, as do the limits (2.17). However the latter limits are preferable in principle, because lower and upper error rates are then both $\rho+O(n^{-1})$. It remains to see if real numerical improvement is made with (2.17); see Section 4.

Further adjustment of $z_\rho^N$ to account for the $n^{-1}$ term in (2.10) is non-trivial, since some of the deviations $\hat{\alpha}_{ab}-\alpha_{ab}$ must be accounted for: $\Pr\{\sqrt{n}\omega(\hat{F}_n,F) \leq \hat{z}_\rho\} \neq \rho+n^{-1}\lambda_2(z) + o(n^{-1})$. In principle one can parallel the development of Hall (1981), but we have not done this.

The preceding results are illustrated in Section 4 in the context of ratio estimation. The next section briefly discusses the standard jackknife and the bootstrap.

## 3. STANDARD JACKKNIFE AND BOOTSTRAP METHODS

### 3.1 Standard Jackknife Expansion

The analysis of Section 2 applies also to the standard jackknife as far as the first-order ($n^{-1/2}$) correction is concerned. We can then define a simple analog of the corrected percentiles (2.16) which may be calculated by standard jackknife methods.

The first step is to observe that the quantities $\tilde{I}_j$ defined in (1.2) may be expressed as (Hinkley, 1982) $\tilde{I}_j = \hat{I}_j - \frac{1}{2}n^{-1}\hat{Q}_{jj} + o_p(n^{-1})$, where $\hat{Q}_{jj} = Q_t(X_j, X_j, \hat{F}_n)$. From this relationship we see that the two variance estimates $\tilde{V}_n$ and $\hat{V}_n$ defined in (1.3) and (1.6) satisfy

$$\frac{n-1}{n}\tilde{V}_n = \hat{V}_n - n^{-1}u(\hat{F}_n) + o_p(n^{-1}) \ ,$$

with $u(F) = E\{I_t(X,F)Q_t(X,X,F)\}$. Therefore, writing $r(F) = u(F)/v(F)$ and then noting that $r(\hat{F}_n) = r(F) + o_p(1)$, we find that

$$\tilde{Z}_n = \frac{\sqrt{n}(T_n-\theta)}{\sqrt{\tilde{V}_n}} = \hat{Z}_n[1 - \tfrac{1}{2}n^{-1}\{r(F) - 1\}]^{-1} + o_p(n^{-1}).$$

Thus the Edgeworth expansion (2.10) applies also to $Z_n$, except for the obvious modification in the $n^{-1}$ term because of the constant factor in square brackets above. In particular, the $n^{-\frac{1}{2}}$ correction term is unchanged and we have

$$\Pr(\tilde{Z}_n \leq z) = \Phi(z) - n^{-\frac{1}{2}}\lambda_1(z)\ \phi(z) + O(n^{-1}). \tag{3.1}$$

By analogy with the discussion in Section 2, corrections to the normal percentile approximations for $\tilde{Z}_n$ can be defined by constructing consistent estimates for $\alpha_{11}$ and $\alpha_{32}$ from components of the standard jackknife procedure. A straightforward analog of (2.16) is

$$\tilde{z}_\rho = z_\rho^N + n^{-1/2}[\tilde{\alpha}_{11} + \frac{1}{6}\tilde{\alpha}_{32}\{(z_\rho^N)^2 - 1\}] \ , \tag{3.2}$$

where

$$(3.3)\qquad \tilde{\alpha}_{11} = \frac{n\tilde{B}_n}{\sqrt{\tilde{V}_n}} - \frac{1}{2\tilde{V}_n^{3/2}} \left(n^{-1}\sum_{j=1}^{n}\tilde{I}_j^3 + 2n^{-2}\sum\sum_{j\neq k}\tilde{I}_j\tilde{I}_k\tilde{Q}_{jk}\right)$$

$$(3.4)\qquad \tilde{\alpha}_{32} = \frac{1}{\tilde{V}_n^{3/2}}\left(-2n^{-1}\Sigma\tilde{I}_j^3 - 3n^{-2}\sum\sum_{j\neq k}\tilde{I}_j\tilde{I}_k\tilde{Q}_{jk}\right) ,$$

with $\tilde{I}_j$ as defined in (1.2), $\tilde{B}_n = -\bar{\tilde{I}}_{\bullet}$ , and

$$\tilde{Q}_{jk} = n\{n\, T_n - (n-1)(T_{n/j} + T_{n/k}) + (n-2)T_{n/(j,k)}\} , \quad j\neq k ,$$

estimating $Q_t(X_j,X_k,F)$ for $j\neq k$ .

In (3.3) we have used the special relationship (Efron, 1982; Hinkley, 1982)

$$\tilde{B}_n = -\bar{\tilde{I}}_{\bullet} = (2n)^{-1}\, E\, Q_t(X,X;F) + O_p(n^{-1}) \quad .$$

It is clear that (3.2) with the first term of $\tilde{\alpha}_{11}$ removed applies to $\sqrt{n}(T_n - \theta - \tilde{B}_n)/\sqrt{\tilde{V}_n}$ , so that the resulting first-order correction to normal percentiles is the same whether or not $T_n$ has the bias adjustment subtracted in the Studentized form. If we write

$$\tilde{\alpha}_{11} = n\tilde{B}_n/\sqrt{\tilde{V}_n} + \tilde{\alpha}_{11}^{*} \quad ,$$

then (3.2) gives corrected equi-tailed $1-2\rho$ confidence limits

$$(3.5)\qquad T_n - \tilde{B}_n \pm \frac{\sqrt{\tilde{V}_n}}{\sqrt{n}}\left(z_{1-\rho}^{N} + n^{-1/2}[\tilde{\alpha}_{11}^{*} + \frac{1}{6}\tilde{\alpha}_{32}\{(z_{1-\rho}^{N})^2 - 1\}]\right) .$$

The correction (3.2) is theoretically justified if $\Pr(\tilde{Z}_n \leq z) = \rho + O(n^{-1})$, which follows from (3.1) if $\tilde{\alpha}_{11} = \alpha_{11} + O_p(n^{-1/2})$ and $\tilde{\alpha}_{32} = \alpha_{32} + O_p(n^{-1/2})$. The latter can be proved quite easily because $\tilde{I}_j = \hat{I}_j + O_p(n^{-1})$, $\tilde{Q}_{jk} = Q_t(X_j, X_k, \hat{F}_n) + O_p(n^{-1})$, and $\alpha_{ab}(\hat{F}_n) = \alpha_{ab}(F) + O_p(n^{-1/2})$. We omit the details.

## 3.2 Bootstrap Method

In Sections 2 and 3.1 we have discussed jackknife methods of approximating distributions, by correcting normal approximations. The closely related non-parametric bootstrap methods of Efron (1982) may be used to approximate the same distributions directly. Recent theoretical work by Beran (1982a,b) suggests that the bootstrap results will be comparable to those obtained by Edgeworth expansion methods, so that the bootstrap should provide a competitive method of non-parametric calculation of confidence limits. (Beran's detailed results do not cover the Studentized forms $\hat{Z}_n$ and $\tilde{Z}_n$ directly.)

Suppose that we are estimating the distribution of

$$Q_n = q(T_n - \theta, \hat{F}_n) \quad ,$$

e.g. $Q_n = \hat{Z}_n$ or simply $Q_n = \sqrt{n}(T_n - \theta)$. The basic bootstrap method works as follows. Construct M independent bootstrap samples $\{X^*_{ij} : j=1,\ldots,n\}$, $i=1,\ldots,M$, by sampling with replacement from the data $(X_1,\ldots,X_n)$. For the i-th sample denote the sample c.d.f. by $\hat{F}^*_n(i)$, the estimate of $\theta$ by $T^*_n(i) = t(\hat{F}^*_n(i))$, and

$$Q^*_n(i) = q(T^*_n(i) - T_n, \hat{F}^*_n(i)) \quad , \qquad i=1,\ldots,M \quad .$$

Then the true distribution

$$G(z) = \Pr(Q_n \leq z)$$

is estimated by

$$(3.6) \qquad \hat{G}(z) = \frac{1}{M} (\text{\# times } Q_n^*(i) \leq z \ , \ i = 1,\ldots,M) \ .$$

The accuracy of $\hat{G}$ will depend on M, and on the degree to which $Q_n$ is pivotal (Chapman, 1983). For example, the Studentized form $\hat{Z}_n = \sqrt{n}(T_n - \theta)/\sqrt{\hat{V}_n}$ should have a more stable distribution than $\sqrt{n}(T_n - \theta)$, with respect to variations in F, so that the bootstrap estimate $\hat{G}$ will tend to be better for the Studentized statistic.

Suppose, then, that the estimate (3.6) is applied with $Q_n = \hat{Z}_n$, and let the estimated percentiles be denoted by

$$\hat{k}_\rho^{BOOT} = G^{-1}(\rho) = [M\rho]^{th} \text{ largest value of } Q_n^*(i).$$

Then approximate equi-tailed $1-2\rho$ confidence limits for $\theta$ are

$$(3.7) \qquad T_n - \hat{k}_{1-\rho}^{BOOT} \sqrt{(\hat{V}_n/n)} \ , \quad T_n - \hat{k}_\rho^{BOOT} \sqrt{(\hat{V}_n/n)} \ .$$

(Note that these differ from limits obtained via $Q_n = \sqrt{n}(T_n - \theta)$.)

The bootstrap method is compared with the earlier jackknife methods in the example of the next section.

## 4. AN EXAMPLE: RATIO ESTIMATION

To illustrate the preceding discussion we look at the relatively simple example of ratio estimation, where $X = (X_1,X_2)$ and the ratio of averages $T_n = \overline{X}_1/\overline{X}_2$ estimates the ratio of means $\theta = E(X_1)/E(X_2)$. We begin by summarizing the theoretical calculations relevant to expansion (2.10) for this problem. Then we discuss numerical illustrations in the context of a particular finite population.

With $X, T_n$ and $\theta$ as just described, define $\beta = E(X_2)$ and

$$\varepsilon(x) = x_1 - \theta x_2 \quad .$$

Then the first three derivatives of $t(F) = \int x_1 dF(x_1,x_2)/\int x_2 dF(x_1,x_2)$ are

$$I_t(x,F) = \beta^{-1}\varepsilon(x), \; Q_t(x,y,F) = -\beta^{-2}\{(y_2-\beta)\varepsilon(x) + (x_2-\beta)\varepsilon(y)\}$$

$$C_t(x,y,z,F) = 2\beta^{-3}\{(y_2-\beta)(z_2-\beta)\varepsilon(x) + (z_2-\beta)(x_2-\beta)\varepsilon(y) + (x_2-\beta)(y_2-\beta)\varepsilon(z)\}.$$

The variance function defined in (2.1) is therefore

$$v(F) = \beta^{-2}\,\sigma_\varepsilon^2 \quad , \qquad \text{with } \sigma_\varepsilon^2 = E\varepsilon^2(X) \quad .$$

Some straightforward but lengthy calculations, using (2.8), (2.9), (2.13) and (2.14), yield the cumulant coefficients

(4.1) $$\alpha_{11} = E\{\tilde{X}_2\tilde{\varepsilon}(X)\} - \tfrac{1}{2}E\,\tilde{\varepsilon}^3(X), \; \alpha_{32} = 6\,E\,\{\tilde{X}_2\tilde{\varepsilon}(X)\} - 2\,E\,\tilde{\varepsilon}^3(X) \; ,$$

$$\alpha_{22} = 3 - 3\,\mathrm{Var}\,(\tilde{X}_2) + \tfrac{7}{4}\{E\;\tilde{\varepsilon}^3(X)\}^2 + 6\,E\,\{(\tilde{X}_2-1)\tilde{\varepsilon}^2(X)\}$$

$$- \; 9\,E\,\{\tilde{X}_2\tilde{\varepsilon}(X)\}E\tilde{\varepsilon}^3(X) + 5[E\{\tilde{X}_2\tilde{\varepsilon}(X)\}]^2 \quad ,$$

and

$$\alpha_{43} = 12 - 12\,\mathrm{Var}\,(\tilde{X}_2) + 12\{E\tilde{\varepsilon}^3(X)\}^2 - 2E\tilde{\varepsilon}^4(X)$$

$$- \; 60E\{\tilde{X}_2\tilde{\varepsilon}(X)\}E\tilde{\varepsilon}^3(X) + 24\,E\,\{(\tilde{X}_2-1)\tilde{\varepsilon}^2(X)\} + 60[E\{\tilde{X}_2\tilde{\varepsilon}(X)\}]^2 \; ,$$

where $\tilde{X}_2 = X_2/\beta$ and $\tilde{\varepsilon}(X) = \varepsilon(X)/\sigma_\varepsilon$ .

The estimates $\hat{\alpha}_{ab} = \alpha_{ab}(\hat{F}_n)$ in (2.16) would be obtained by substituting sample moments in the above formulae, with $\hat{\varepsilon}(X_j) = X_{1j} - T_n X_{2j}$ in place of $\varepsilon(X_j)$.

In a typical application the variance of $\varepsilon(X)$ conditional on $X_2$ would not be constant, so that $E\{(X_2-\beta)\varepsilon^2(X)\}$ would not be zero; nor would be $E\varepsilon^4(X)$. Typical <u>models</u> would predict $E\{X_2\varepsilon(X)\} = E\varepsilon^3(X) = 0$, making $\alpha_{11} = \alpha_{32} = 0$, but in many real populations (including the one below) this is not the case.

Our example population is a set of 307 pairs, $X_1$ = distance of bus trip and $X_2$ = ticket revenue. The data are graphed in Figure 1. Calculation reveals that $\theta = 0.081694$, $\beta = 153.5$, $\mathrm{Var}(X_2) = 19900$ ,

$$\sigma_\varepsilon^2 = E\varepsilon^2(X) = 20.01\ , \quad E\varepsilon^3(X) = -113.5\ , \quad E\varepsilon^4(X) = 4217,$$

$$E\{X_2\varepsilon(X)\} = -81.65\ , \quad E\{(X_2-\beta)\varepsilon^2(X)\} = 6015.$$

Given these, equation (4.1) produces the values

$$\alpha_{11} = 0.515 \qquad \alpha_{32} = 1.822$$

$$\alpha_{22} = 13.66 \qquad \alpha_{43} = 38.66$$

for the coefficients in (2.10)-(2.12).

An initial series of computations was carried out to compare the "exact" distributions of $\hat{Z}_n$ and $\tilde{Z}_n$ with expansion (2.10); recall that only the first correction term is valid theoretically for $\tilde{Z}_n$. "Exact" distributions are based on 10,000 pseudo-random samples of size $n$ from the population. Figure 2a graphs these exact distributions for n=50 on the probit scale, together with one- and two-term corrections to the standard normal approximation. Figure 2b compares the exact distribution of $\hat{Z}_n$ with the average of bootstrap estimate $\hat{G}(z)$ defined in (3.6), this average obtained from 1000 data samples with M=1000 bootstrap subsamples per data sample. We conclude from these numerical results that: (i) the normal approximation for $\hat{Z}_n$ is inadequate at n=50; (ii) the one-term Edgeworth correction does not give satisfactory improvement, whereas the two-term Edgeworth correction does; (iii) the bootstrap produces an excellent approximation <u>on average</u>.

The emphasis in the last remark is necessary because bootstrap estimates $\hat{G}(z)$ vary from sample to sample, as would sample estimates of Edgeworth expansions. As an illustration of this, again for n=50, Table 1 gives the

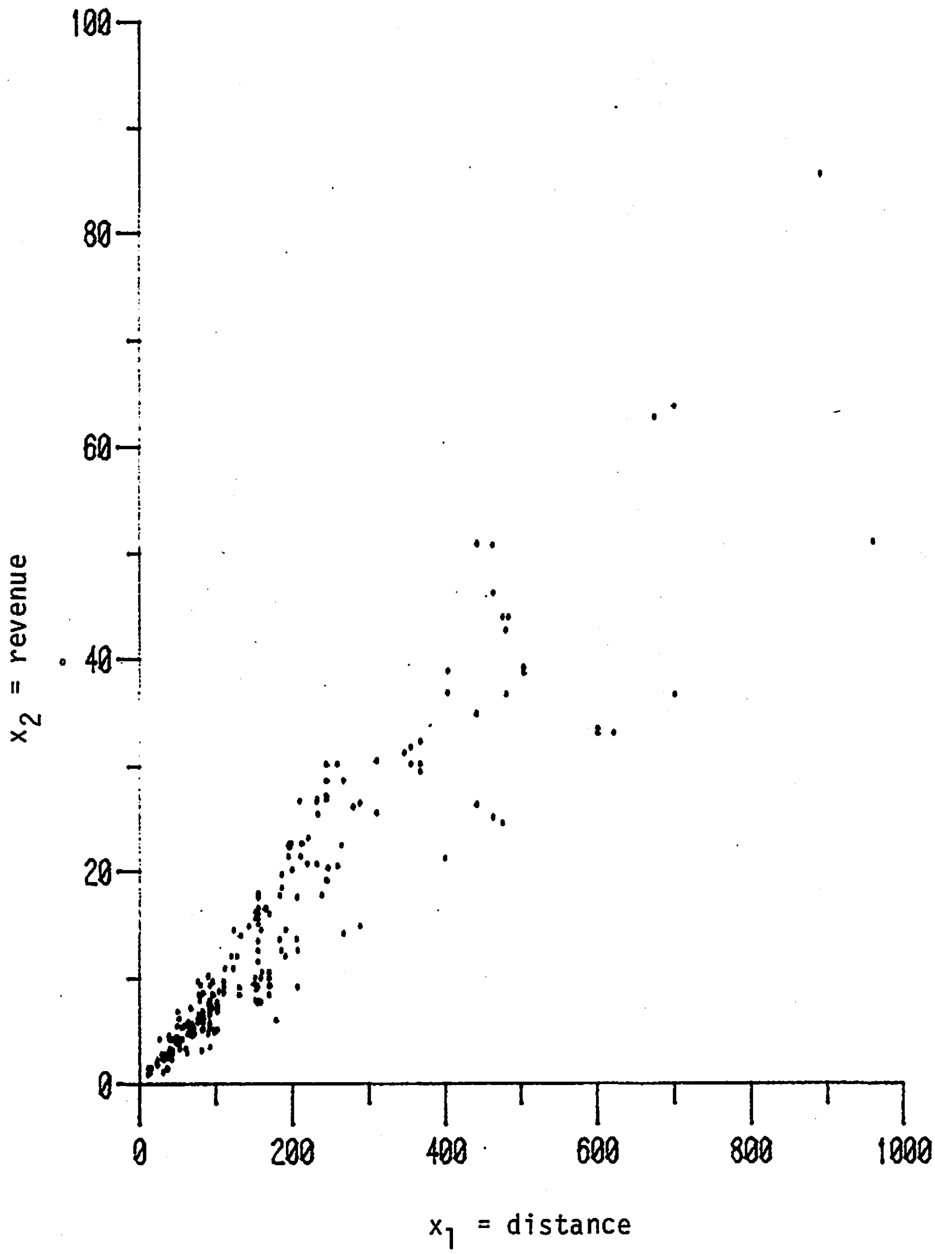


FIGURE 1. Scatterplot of bus ticket data : 307 pairs

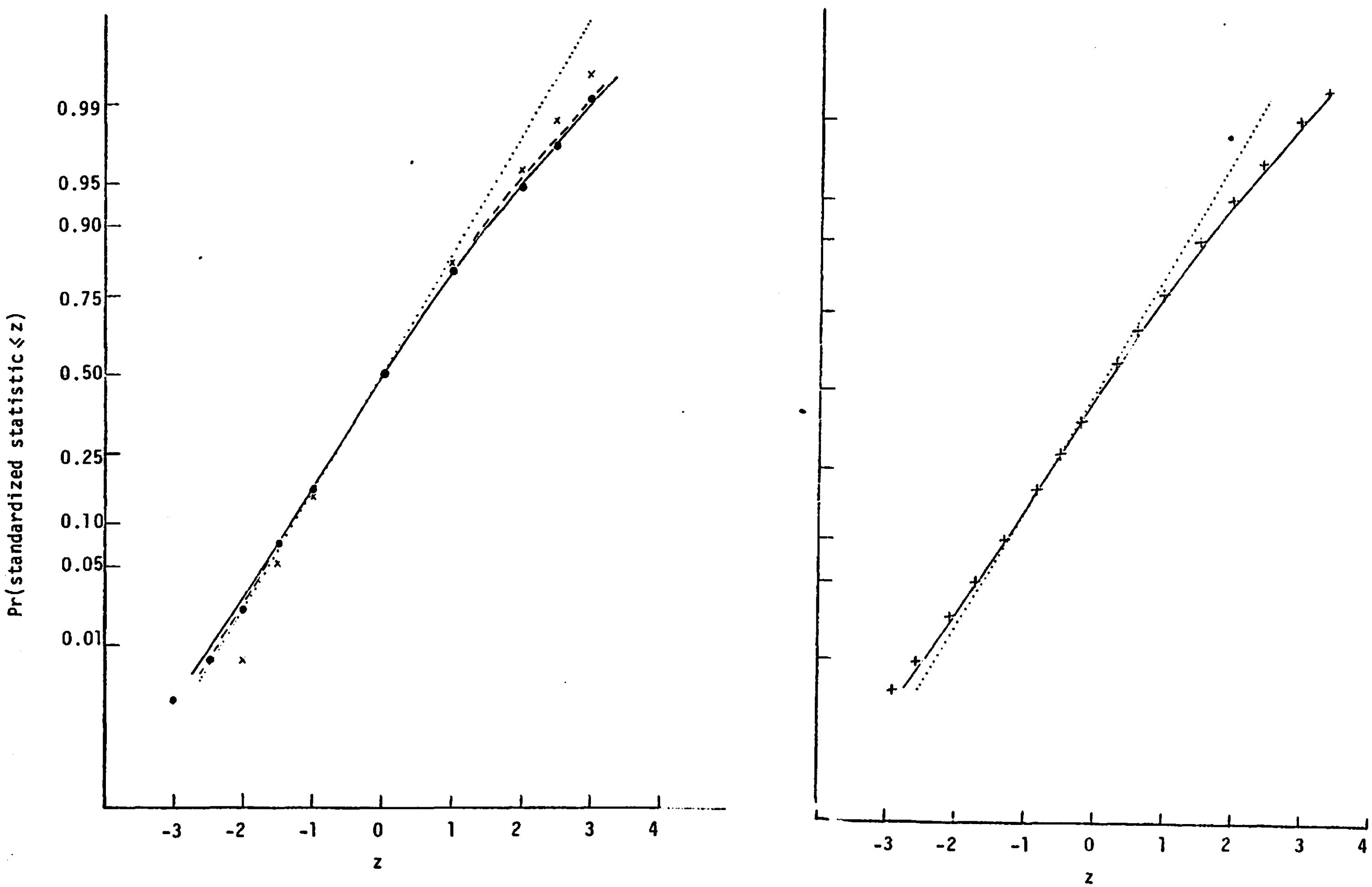


FIGURE 2. Exact distributions of $\tilde{Z}$ and $\hat{Z}$, Edgeworth approximations and bootstrap estimates for $\hat{Z}$, n=50.

Key: .... normal approx.; --- dist. of Z; ——— dist. of Z; x one-term Edgeworth; ● two-term Edgeworth; + mean of 1000 bootstrap estimates each with M=1000.

averages and standard deviations of $\hat{G}(z)$ and of one- and two-term Edgeworth corrections using $\alpha_{ab} = \alpha_{ab}(\hat{F}_n)$ to estimate $\alpha_{ab}$ .

TABLE 1. Means and standard deviations of bootstrap and Edgeworth expansion estimates of $Pr(\hat{Z}_n \leq z)$ for n=50. (Obtained from 1000 pseudo-random samples.)

| z= | | -3 | -2 | -1.5 | -1 | 0 | 1 | 1.5 | 2 | 3 |
|---|---|---|---|---|---|---|---|---|---|---|
| true probability | | .002 | .028 | .07 | .160 | .49 | .82 | .90 | .95 | .99 |
| bootstrap | mean | .004 | .030 | .07 | .16 | .49 | .82 | .91 | .95 | .99 |
| | st. dev. | .003 | .009 | .01 | .02 | .02 | .02 | .02 | .02 | .01 |
| one-term Edgeworth | mean | .001 | .018 | .06 | .15 | .49 | .83 | .93 | .97 | 1.00 |
| | st. dev. | .001 | .007 | .01 | .01 | .01 | .01 | .01 | .01 | .001 |
| two-term Edgeworth | mean | .001 | .025 | .08 | .18 | .49 | .80 | .91 | .96 | 1.00 |
| | st. dev. | .004 | .009 | .01 | .02 | .01 | .04 | .02 | .01 | .004 |

The final numerical results are empirical evaluations of the several confidence limit methods described in Sections 2, 3 and 4. Still with n=50, Table 2 shows (estimated) left and righterror rates for nominal equi-tailed 90% and 95% confidence intervals on the ratio $\theta$. (A left error occurs when $\theta$ is to the left of the confidence interval, with right error analogously defined.) As earlier numerical results might suggest, one-term Edgeworth corrections to percentiles do not improve confidence limits very much in this particular problem. The simpler bootstrap method based on $\hat{Z}_n$ works better.

TABLE 2. Error rates of jackknife confidence limit methods based on $\hat{Z}_n$ and $\tilde{Z}_n$ when n=50. Nominal equitailed error rates 10% and 5%.

| Method | nominal left and right error rates % | *empirical error rates left | right |
|---|---|---|---|
| infinitesimal jackknife | 5 | 8½ | 5½ |
| with normal percentiles | 2½ | 5 | 2½ |
| infinitesimal jackknife | 5 | 8 | 5 |
| with percentiles (2.16) | 2½ | 5 | 2½ |
| standard jackknife | 5 | 7½ | 5 |
| with percentiles (3.2) | 2½ | 5 | 2½ |
| bootstrap percentiles | 5 | 5 | 4 |
| for Z (M= 999) | 2½ | 3½ | 1½ |

* based on 10,000 pseudo-random samples except for bootstrap method, which is based on 1,000 pseudo-random samples.

## 5. DISCUSSION

Edgeworth expansions present a natural way of correcting inaccuracies in normal approximations. We have shown how to do this when estimates are Studentized using jackknife techniques. The resulting corrections are, in general, complicated and seem to be equalled by direct bootstrap estimates of distributions.

One-term corrections to percentiles will not be satisfactory if fourth-moment effects are large, as in our example. Two-term corrections to percentiles are, in principle, obtainable, but their complexity would be such as to make the bootstrap a more appealing method of calculating percentiles for the Studentized estimate. For more complete discussions of bootstrap confidence limit methods see Efron (1982), Hinkley (1982) and Chapman (1983).

ACKNOWLEDGEMENTS:

Special thanks go to Philip Chapman for discussions on the bootstrap, and for carrying out the numerical computations. We are grateful to Frank Martin for providing the data population.

## APPENDIX: Summary of Calculus Operations for Functionals

We record here several of the operational results concerning von Mises derivatives. These are proved directly by applying the definition (2.1). For simplicity we write $\dot{a}_x$ for the von Mises derivative of $a(F)$ at $(x,F)$.

$$\text{if } \quad t(F) = \Sigma c_i t_i(F) \text{ , then } \quad \dot{t}_x(F) = \Sigma c_i \dot{t}_{ix}(F)$$

$$\text{if } \quad t(F) = t_1(F)t_2(F) \text{ , then } \quad \dot{t}_x(F) = t_1\dot{t}_{2x} + \dot{t}_{1x}t_2$$

$$\text{if } \quad t(F) = \{t_1(F)\}^{-1} \text{ , then } \quad \dot{t}_x = -\frac{1}{t_1^2}\dot{t}_{1x}$$

$$\text{if } \quad t(F) = h(t_1(F),\ldots,t_k(F)) \text{ , then } \quad \dot{t}_x = \Sigma\frac{\partial h}{\partial t_j} \cdot \dot{t}_{jx}$$

$$\text{if } \quad t(F) = \int a(u,H)dF(u) \text{ where } H \text{ is a c.d.f. not depending on } F,$$

$$\text{then } \quad \dot{t}_x = a(x,H)-t$$

$$\text{if } \quad t(F) = \int a(u,F)dF(u) \text{ , then } \quad \dot{t}_x = \int \dot{a}_x(u,F)dF(u) + a(x,F) - t \; .$$

Results for second and higher derivatives are obtained by using the above in conjunction with definitions such as (2.6).

<u>Example</u> If $v(F) = \int I_t^2(u,F)dF(u)$ , then

$$\begin{aligned} I_v(x,F) = \dot{v}_x &= \int 2I_t(u,F)\{I_t(u,F)\}_x^{\cdot} \, dF(u) + I_t^2(x,F) - v \\ &= 2\int I_t(u,F)\{Q_t(u,x)F) - I_t(x,F)\}dF(u) + I_t^2(x,F) - v \\ &= 2E\{I_t(U,F)Q_t(U,x,F)\} + I_t^2(x,F) - v \; . \end{aligned}$$

# REFERENCES

BERAN, R.J. (1982a). Estimated sampling distributions: the bootstrap and competitors. Ann. Statist. 10, 212-225.

BERAN, R.J. (1982b). Jackknife approximations to bootstrap estimates. Unpublished report, Department of Statistics, Univ. of Calif., Berkeley.

BHATTACHARYA, R.N. and GHOSH, J.K. (1978). On the validity of the formal Edgeworth expansion. Ann. Statist. 6, 433-451. [Corrigendum ibid 8, 1399.]

CHAPMAN, P. (1983). Bootstrap methods. Ph.D. thesis (in preparation). University of Minnesota, School of Statistics.

EFRON, B. (1982). The Jackknife, the Bootstrap and Other Resampling Plans. SIAM-CBMS Monograph No. 38, Philadelphia: SIAM.

HALL, P. (1983). Inverting an Edgeworth expansion. Ann. Statist. 11 (June issue).

HINKLEY, D.V. (1982). Subsampling Methods: Jackknife, Bootstrap and Cross-validation. Draft version of monograph to be published by Chapman and Hall in late 1983.

REEDS, J.A. (1976). On the Definition of von Mises Functionals. Ph.D. thesis, Harvard University.

WITHERS, C.S. (1983). Expansions for the distribution and quantiles of a regular functional of the empirical distribution with applications to nonparametric confidence intervals. Ann. Statist. 11, to appear.